3.0 FURTHER TOPICS IN BOOTSTRAPPING

3.1 Introduction

There is much more to bootstrapping than can be covered here. The book by Efron and Tibishirani (1993) provides a detailed reference. Some useful further features are given here, and additional examples appear in subsequent chapters.

3.2 Estimates of bias

Bootstrapping provides a handy way to check for bias in an estimate. Suppose we calculate some statistic, such as the variance, from a set of data. We would like that statistic to provide an unbiased estimate of the true parameter, σ^2. If we represent the statistic that we intend to calculate from the data as $\hat{\theta} = s(\mathbf{x})$, where $\mathbf{x}$ is the vector of observed data, $x_1, x_2, ..., x_n$ and $s(\mathbf{x})$ is some function of that data, then we would like to have the expected value of that function to be $E(\hat{\theta}) = E[s(\mathbf{x})] = \theta = \sigma^2$. When the theoretical distribution of the estimator is known or assumed to be known, then an unbiased estimator can often by found by the methods of mathematical statistics, i.e., we can find the expected value of a trial statistic directly. Thus we know that the sample variance:

$s^2 = \dfrac{\Sigma(x_i - \bar{x})^2}{n-1}$ has expected value $E(s^2) = \sigma^2$ for the normal distribution of eq. (1.3). In this case, $\hat{\theta} = s(\mathbf{x}) = s^2$ (the notation can be a little confusing, as we use the notation $s(\mathbf{x})$ to represent any statistic calculated from a data set, $\mathbf{x}$, whereas s is also commonly used to represent a specific quantity, the sample estimate of the standard deviation).

Very often we are not sure what theoretical distribution may be appropriate for an observed sample, and it is frequently true that there may not be any such distribution. Statisticians thus spend a lot of time trying to choose the "right" distribution or manipulating (transforming) the data to approximate some known distribution. Bootstrapping can avoid a lot of that trouble and uncertainty. In this section we consider how bootstrapping can be used to check for bias in an estimator. We define bias as:

$$\text{bias}_F = E_F[s(\mathbf{x})] - t(F) \qquad (3.1)$$

where the subscript F serves as a reminder that the bias and expectation are taken with respect to some probability distribution function F (quite likely an unknown distribution), and $\theta = t(F)$ denotes our statistic as calculated from the true probability distribution. The bootstrap estimate of bias is calculated as:

$$\widehat{\text{bias}}_B = \hat{\theta}^*(\bullet) - t(\hat{F}) \qquad (3.2)$$

where $\hat{\theta}^*(\bullet)$ is the mean of our estimator calculated from a large number of bootstrap samples, and $t(\hat{F})$ is the same estimator calculated from the original data.

Example 3.1. A numerical example from a normal distribution may help fix ideas at this point. The following 10 "observations" are from a normal distribution with mean zero:

1.6718,-3.061,0.9338,2.8766,-1.248,2.6206,0.3212,-1.121,0.0475,1.7129

Consider estimating the variance of these observations from:

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n} \tag{3.3}$$

This is the equation for variance often used by engineers, and produced by some of the earlier pocket calculators. If we apply the formula to the 10 original observations it gives $t(\hat{F}) = 3.177$. Now calculate the average of 2,000 bootstraps using this estimator. This gives $\hat{\theta}^*(\bullet) = 2.799$, and we get:

$$\widehat{bias}_B = \hat{\theta}^*(\bullet) - t(\hat{F}) = 2.799 - 3.177 = -0.378,$$

a negative bias, suggesting that we have an underestimate. From theory, we know that an unbiased estimator comes from

$$s^2 = \frac{S(x_i - \bar{x})^2}{n-1}$$

which gives us $s^2 = 3.530$ from the above data set. Our bias estimate is negative, meaning that we underestimate the true quantity. We could thus add this quantity, 0.378, to our underestimate from the original data, $t(\hat{F}) = 3.177$, getting an improved estimate (3.555). It is perhaps better to define a bias-corrected estimator :

$$\theta'' = \hat{\theta} + [\hat{\theta} - \hat{\theta}^*(\bullet)] = 2\hat{\theta} - \hat{\theta}^*(\bullet) = 2(3.177) - 2.799 = 3.555 \tag{3.4}$$

which gives the same result. This is close to the result (3.530) one would get by using the proper equation $s^2 = \frac{S(x_i - \bar{x})^2}{n-1}$ in the first place. The point here is that we often don't know what the proper equation is, and the bootstrap provides a way to check for bias in whatever equation we do have available to estimate some quantity. Eq. (3.4) came very close to the correct answer in this example, but in practice, if we have indications of an important bias, simply correcting by eq.(3.4) may not necessarily improve the situation. The estimator may be subject to a great deal of variability, so that the adjustment may not help. The essential conclusion here, is that if the bootstrap indicates small bias and small standard error, then we can be very comfortable indeed with our estimator, even if we don't have a theoretical model. Note that this result came out very close to the expected answer just by chance; repeating it gives a smaller bias, as will likely be evident in Exercise 3.10.1.

3.3 An improved bias estimate

Efron and Tibishirani (1993:Ch. 10) recommend an improved bias estimate that converges on the asymptotic estimate with a smaller bootstrap sample, B. They define a resampling vector for each bootstrap sample that contains the proportions of that bootstrap sample calculated from the

frequencies with which the individual entries are observed. Thus in Fig. 2.1, the original data set was:

1	2	3	4	5	6	7	8	9	10
13	106	203	131	160	8	67	61	11	301

and the resampling vector for the first bootstrap sample of Fig. 2.1 would be:

0	0.4	0.2	0	0.2	0.1	0	0	0	0.1

that is, 13 doesn't appear at all, but 106 appears 4 times, and so on. They then average these B resampling vectors, obtaining a final vector with the proportions averaged over all B vectors, denoted as $\bar{\mathbf{P}}^*$ and use this instead of $t(\hat{F})$ in eq.(3.2), obtaining,

$$\overline{bias}_B = \hat{\theta}^*(\bullet) - T(\bar{\mathbf{P}}^*) \qquad (3.5)$$

Example 3.2 We can illustrate the improved bias correction by using the original data of Example 3.1, but using B = 500. The proportions of $\bar{\mathbf{P}}^*$ will add to unity and $T(\bar{\mathbf{P}}^*)$ is then calculated as a weighted variance, using the proportions of $\bar{\mathbf{P}}^*$ as weights(w_i). The weighted mean is $\bar{x}_w = \Sigma w_i x_i$, and the variance is calculated as $s^2 = \frac{\Sigma w_i (x_i - \bar{x}_w)^2}{n}$. A run with B = 500 gave $\overline{bias}_B = \hat{\theta}^*(\bullet) - T(\bar{\mathbf{P}}^*) = 2.910 - 3.223 = -0.313$ and bias-corrected estimate of s^2 is then 3.536, which is very close to the result (3.530) obtained by dividing by n-1, as should be done in practice. I would be inclined to use this approach on complicated problems, where bootstrapping uses a fair bit of computer time. Otherwise, one can simply use a sizable number (say 2,000) of bootstraps as in Example 3.1, inasmuch as it is likely that percentile confidence limits will be also be calculated in a practical example--here we know the “right” answer (i.e., divide by n-1) from theory.

3.4 Cross-validation

Models applied to ecological data may serve various purposes, but the more important uses may be to see how well we understand the data, and to make predictions. One of the earliest approaches to evaluating predictions from a model is very simple. One develops and fits the model on half of the available data, and then tests its predictions on the other half. Using all of the data for development and testing invariably results in underestimating the prediction errors. With the increased computing power now available, models can be fit to various subsets of the data and tested on the remainder. The logical outcome appears to be fitting the model to all but one of the observations, making a prediction for the remaining observation and repeating the process for all n observations, getting n predictions and deviations from predicted value. The variance is then calculated as:

$$CVE = \frac{1}{n} \sum (y_i - \hat{y}_{-i})^2 \qquad (3.6)$$

where $\hat{y}_{-i}$ denotes a predicted value based on all of the observations except the i^{th} value, and the summation runs from 1 to n. One thus deletes one observation, fits the model, and makes a prediction for the missing value, doing this n times to calculate the cross-validation error, CVE.

Example 3.3 Cross-validation error. To demonstrate the cross-validation idea, we use a larger set of regression data (n = 30). The data are as follows:

No.	x	y	No.	x	y
1	14.48	18.30	16	10.00	17.60
2	22.49	19.98	17	20.06	19.33
3	19.71	18.51	18	9.70	17.89
4	29.89	21.00	19	26.86	20.79
5	30.01	21.00	20	34.23	20.36
6	21.61	19.77	21	27.53	20.65
7	16.71	18.77	22	19.88	18.82
8	26.78	20.26	23	21.99	19.26
9	17.85	18.70	24	21.09	20.02
10	33.04	20.25	25	28.68	20.38
11	18.92	18.66	26	21.91	19.61
12	20.23	18.45	27	28.50	20.45
13	28.24	20.60	28	20.01	18.93
14	29.77	21.10	29	17.62	19.09
15	22.92	19.24	30	22.90	19.87

Applying cross-validation is simple in this case, with the only difficulty being one of arranging to drop each observation in turn We then can compute CVE from eq.(3.6), which turns out to be 0.201. Note that the values of $\hat{y}_{-i}$ in eq. (3.6) are computed from individual regressions dropping the i^{th} point, and y_i is the y-value of the i^{th} observation. For comparison the value of the variance about regression (eq.(2.5)) is 0.182. This is somewhat smaller, as might be expected because the deviations from regression are from a normal distribution in this example, and thus the normal-theory model gives the best estimate. Cross-validation would be used only in the absence of suitable theoretical estimators for the model parameters.

3.5. Bootstrapping for predictions. According to Efron and Tibishirani (1993:Chap. 17) bootstrapping offers an alternative to cross-validation. They focus on estimating the variance (in this example, variance about regression as given by eq. (2.5)). There are two stages in the bootstrapping approach. The first is to obtain bootstrap samples from the data set (given in Example 3.3), calculate regression lines for each bootstrap sample, and calculate a variance about each such regression (eq.(2.5)) using the original data set as x_I and y_i

values. The second stage is to calculate the variance about regression for the bootstrap sample now using only the bootstrap sample (the y_i* values). Thus, two variances about regression are calculated from the same regression equation, using the following variances (the regression coefficients are those calculated on the bootstrap data in both cases):

$$s^2 = \frac{\Sigma(y_i - (a + bx_i))^2}{n-2} \quad \text{and} \quad s^2 = \frac{\Sigma(y_i^* - (a + bx_i^*))^2}{n - 2}$$

Often, the variance about regression obtained from the bootstrap sample will be appreciably smaller than that obtained from the original data; and it is the mean difference of these two variances that is sought here. A few values from bootstrapping follow:

Bootstrap no.	Variance about regression using orig. data for y and x	Variance about regression using bootstrap values of y and x	Difference
1	0.1786	0.1905	-0.0119
2	0.1804	0.1863	-0.0059
3	0.1820	0.1519	0.0301
4	0.1724	0.1514	0.0210
5	0.1814	0.1759	0.0055
6	0.1948	0.1301	0.0647
7	0.1863	0.1277	0.0586
8	0.2023	0.1449	0.0574
9	0.1731	0.1189	0.0542

The "inflation factor" (mean difference) is added to the variance of the original data set to give an improved estimate. In the present example, the mean difference in the two variances about regression is small (0.028) so adding it to the variance about regression calculated from the original data (0.184) makes only a minor change. However adding the correction to 0.184 gives a value (0.212) closer to that obtained in example 3.3. Nonetheless the best estimate is that of the original regression calculation because the data of Example 3.3 were generated from a bivariate normal distribution. Note that the bootstrap operation is a sampling procedure so that there will be small differences in the mean differences in repeat runs. Two further runs with B=2000 gave mean differences of 0.027 and 0.029.

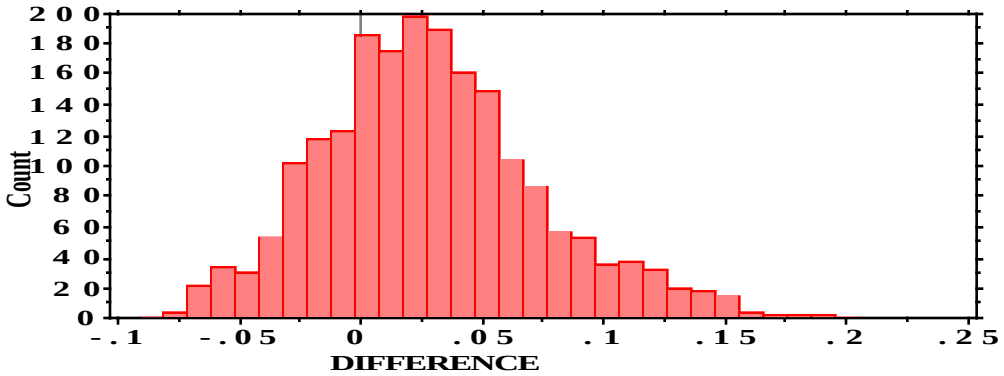

Fig. 3.1. Mean differences between variances about regression calculated from 2000 bootstraps. The differences are between variances about regression calculated from a regression line based on an individual bootstrap sample, with the first variance calculated using the original data, and the second calculated from the data of the particular bootstrap sample.

It turns out that the correction indicated above is really a correction for bias. Inasmuch as the data we used in this example were normally distributed, the variance estimate should be unbiased, and the bootstrap analysis consequently comes up with a minor change, as would be expected with an unbiased estimator. The data used were drawn from a bivariate normal distribution, which is the basis for the normal theory confidence limits on a correlation coefficient, so it is worthwhile to compare (Fig. 3.2) the confidence limits on r based on nonparametric bootstrapping with the normal theory limits in this example. The two different calculations of confidence limits are now appreciably closer than they were in Example 2.4. However, regression data approximating the bivariate normal distribution are not often encountered in practice, because one usually somehow selects the x-values used, rather than obtaining them at random, as the bivariate normal regression theory assumes.

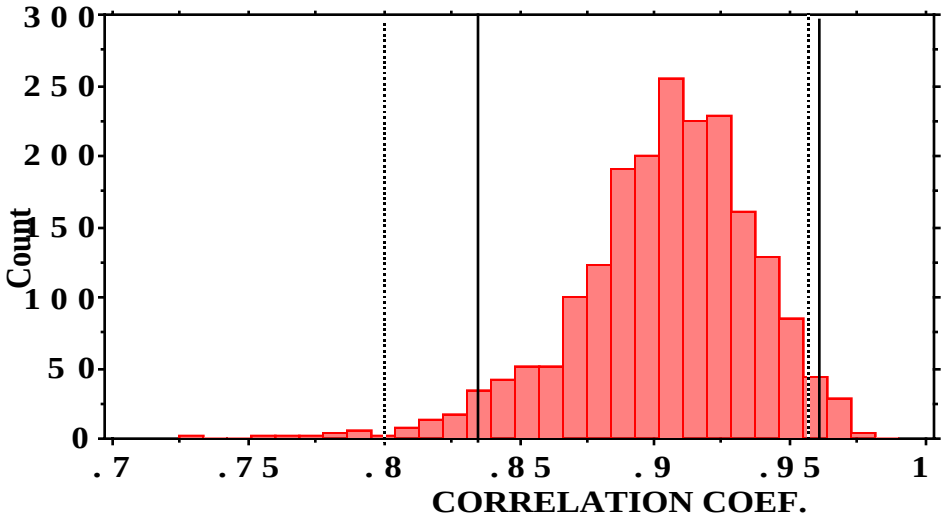

Fig. 3.2. Distribution of 2,000 bootstrap samples for correlation coefficient obtained from the data of Example 3.3. Solid lines show approximate 95% confidence limits based on bootstrapping, while dotted lines show 95% limits based on normal theory (see Example 2.4).

3.6 Improved confidence intervals

Efron and Tibishirani (1993:Chap. 14) recommend an improved bootstrap confidence interval which they call the BCa method ("bias-corrected and accelerated"). These intervals require more calculations than the simple percentile bootstrap confidence intervals. The data from Example 3.3 will be used to illustrate. Two proportions, α_1 and α_2 are calculated, and the total number of bootstraps, B, multiplied by these values. One then orders the B bootstrap values and takes the $\alpha_1 B$ and $\alpha_2 B$ values as confidence limits. Thus if B = 2,000 and α_2 turns out to be 0.0166 as in the example here, then we take the bootstrap value 0.0166(2000) = 33 as the ordered value of the bootstrap replications that gives the lower confidence limit.

Two initial values are calculated. The first is a bias-correction:

$$\hat{z}_o = \Phi^{-1}\left(\frac{\#\{\hat{\theta}^*(b)<\hat{\theta}\}}{B}\right) \qquad (3.7)$$

here, $\Phi^{-1}(\ )$ indicates the inverse of the cumulative normal distribution, which can be looked up in tables and is available in various computer programs (in Microsoft Excel it is "NORMSINV)". The quantity in parentheses has, as numerator, the number of bootstrap samples that are less than the parameter estimate. In the example given below, we consider the correlation coefficients of Section 3.5, for which the correlation coefficient calculated from the original data was r = 0.9043, so we tally the number of bootstrap results that were less than this value and divide by B = 2,000, and look up the inverse cumulative normal value. In this case (Fig. 3.2) there were 942 values less than r = 0.9043, so we look up 942/2000 = 0.471 in the inverse normal tables, getting –0.0728 for $\hat{z}_o$.

The second value is the "acceleration", $\hat{a}$. This is calculated by jackknifing, (Section 3.7) using much the same procedure as in the cross-validation example above. We delete each observation in the original data set in turn, and calculate the correlation coefficient from the remaining observations. From the 30 observations tabulated in Example 3.3, we thus get 30 correlation coefficients, which are here designated as $\hat{\theta}_{(i)}$, where the subscript (i) indicates that the parameter estimate $\hat{\theta}$ (here, r) has been calculated from the original data set with each observation deleted in turn, and $\hat{\theta}_{(.)}$ is used to indicate the average of these 30 values. The estimate of "acceleration" is then calculated as:

$$\hat{a} = \frac{\Sigma(\theta_{(.)}-\theta_{(i)})^3}{6\{\Sigma(\theta_{(.)}-\theta_{(i)})^2\}^{3/2}} \qquad (3.8)$$

where the summations are from 1 to n = 30 in this example. These two parameters ($\hat{z}_o, \hat{a}$) are then used to calculate α_1 and α_2, but require two more values for the calculation. These are designated $z^{(\alpha)}$ and $z^{(1-\alpha)}$, and are the values that cut off a proportion, α, from each tail of the unit normal

distribution.. For 95% confidence limits, we look up $z^{(\alpha)}$ as 0.975 in the inverse cumulative normal table, getting -1.95996, and use 0.025 for $z^{(1-\alpha)}$, giving -1.95996. The calculations then are:

$$\alpha_1 = \Phi\left[\hat{z}_o + \frac{\hat{z}_o + z^{(\alpha)}}{1 - \hat{a}(\hat{z}_o + z^{(\alpha)})} \right] \qquad (3.9)$$

and

$$\alpha_2 = \Phi\left[\hat{z}_o + \frac{\hat{z}_o + z^{(1-\alpha)}}{1 - \hat{a}(\hat{z}_o + z^{(1-\alpha)})} \right]$$

where values of $\Phi(\ )$ are to be looked up in tables of the cumulative normal distribution (in Excel, these are available as NORMSDIST). As noted in the introduction to this section, the lower limit is the 33d ordered value of the bootstrapped correlation coefficients (0.825), and the upper limit is $\alpha_1 B$ = 0.964(2000) giving the 1927th ordered value (0.957).

The calculations are a little onerous to produce the first time, but if they are set up in a spreadsheet, then calculations for a new set of confidence limits only take a few changes. One is to insert the new set of jackknife values in the first column and the second is to change the fraction in eq. (3.7). If the sample size is different, one needs to expand or contract the spreadsheet. An example for the data of Example 3.3 is given below.

The improvement in bootstrap confidence limits in this example is not large, but suggests that the calculations do result in better bootstrap confidence limits. Data for Example 3.3 came from a bivariate normal distribution in which ρ = 0.90. From normal theory, 95% confidence limits were 0.803 to 0.954, while the percentile bootstrap limits in one run with B = 2,000 were 0.830 to 0.959. The improved confidence limits were 0.825 to 0.957, giving a lower limit closer to the normal theory result. Percentile limits vary a little in successive runs, giving 0.837-0.958 and 0.833-0.960 in two additional runs with B = 2,000.

	Correlations	CUBE TERM	SQ TERM
1	0.8978444	2.67703E-07	4.1537E-05
2	0.9075823	-3.57065E-08	1.08434E-05
3	0.9083542	-6.71633E-08	1.6523E-05
4	0.8996521	9.97194E-08	2.1504E-05
5	0.8993546	1.2017E-07	2.43518E-05
6	0.9066940	-1.39052E-08	5.78254E-06
7	0.9018049	1.53344E-08	6.17225E-06
8	0.9027515	3.63677E-09	2.36489E-06
9	0.9018945	1.37354E-08	5.73536E-06
10	0.9181345	-2.65399E-06	0.00019169
11	0.9038609	7.86151E-11	1.83515E-07
12	0.9118957	-4.40089E-07	5.78576E-05
13	0.9013883	2.4416E-08	8.41621E-06
14	0.9008332	4.12818E-08	1.19446E-05
15	0.9065879	-1.2145E-08	5.28363E-06

16	0.8874195	4.80101E-06	0.000284591	
17	0.9044161	-2.03529E-12	1.60602E-08	
18	0.8932690	1.3384E-06	0.000121448	
19	0.9068001	-1.58286E-08	6.30419E-06	
20	0.9205568	-4.30489E-06	0.000264632	
21	0.9029564	2.36797E-09	1.77658E-06	
22	0.9045651	-2.09626E-11	7.60263E-08	
23	0.9048018	-1.34558E-10	2.62587E-07	
24	0.9130072	-6.62566E-07	7.6001E-05	
25	0.9019668	1.2528E-08	5.39412E-06	
26	0.9047293	-8.51567E-11	1.93559E-07	
27	0.9014553	2.27629E-08	8.03189E-06	
28	0.9040066	2.26062E-11	7.995E-08	
29	0.9046723	-5.61539E-11	1.4664E-07	
30	0.9054246	-1.46324E-09	1.28887E-06	
AVERAGE	0.9042893	-1.44489E-06	0.00118043	SUMS

Z(0)-HAT	-0.07276	
ACCELERATION	-0.0059377	
Z(0)-HAT+Z-ALPHA	1.88716	1.95996
1-ACCEL(Z(0)-HAT+Z-ALPHA)	1.0112055	
RATIO	1.8662478	
ALPHA1	0.9635494	1927.0988
Z(0)-HAT+Z(1-ALPHA)	-2.03276	-1.95996
1-ACCEL(Z(0)-HAT+Z(1-ALPHA)	0.987930	
RATIO	-2.057595	
ALPHA2	0.016569	33.139

3.7 The jackknife

The jackknife technique, as noted in the introduction to Chapter 2, predates bootstrapping, and was originally derived (Quenouille(1956)) to evaluate biases in an estimator. The technique is very simple and easy to apply. Given an original data set, one simply leaves out each observation in turn and calculates the statistic of interest on the remaining observations, as was done in the calculations for improved confidence limits above, getting $\hat{\theta}_{(i)} = s(\mathbf{x}_{(i)})$, where $\mathbf{x}_{(i)}$ is the vector of observations with the i^{th} observation removed, and s() denotes some statistic calculated from these observations. The bias estimate is calculated as:

$$\hat{\text{bias}}_{jack} = (n-1)(\hat{\theta}_{(\cdot)} - \hat{\theta}) \qquad (3.10)$$

where $\hat{\theta}_{(\cdot)}$ denotes the mean of the $\hat{\theta}_{(i)}$ and $\hat{\theta}$ is the statistic estimated from the original data. The jackknife estimate of standard error is:

$$\hat{\text{se}}_{jack} = [\frac{n-1}{n} \Sigma (\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)})^2]^{1/2}. \qquad (3.11)$$

We can illustrate the calculations with the data of Example 2.1. The following table shows the 10 original observations and the 10 jackknife

samples created by dropping each observation in turn. If we consider the mean as the statistic to be jackknifed, then the bias estimate from eq.(3.10) turns out to be zero, inasmuch as the mean of the original observations necessarily equals the grand mean of the jackknife samples.

Orig.	Jackknife samples									
data	1	2	3	4	5	6	7	8	9	10
13		13	13	13	13	13	13	13	13	13
106	106		106	106	106	106	106	106	106	106
203	203	203		203	203	203	203	203	203	203
131	131	131	131		131	131	131	131	131	131
160	160	160	160	160		160	160	160	160	160
8	8	8	8	8	8		8	8	8	8
67	67	67	67	67	67	67		67	67	67
61	61	61	61	61	61	61	61		61	61
11	11	11	11	11	11	11	11	11		11
301	301	301	301	301	301	301	301	301	301	
Ave.	116.4	106.1	95.33	103.3	100.1	117	110.4	111.1	116.7	84.44

Eq. (3.11) gives $\hat{se}_{jack}$ = 30.15, while a bootstrap estimate of standard error [eq.(2.1)] is 28.70, and the standard error of the original data is also 30.15, as it should be in this case, because the jackknife standard error formula gives the same result for the standard error of a mean.

Example 3.4 Jackknifing a regression equation

Grizzly bears are very difficult to census due to the fact that they tend to stay in heavy cover when food conditions are good and are thus not visible from the air. Also, they range very widely, are difficult (and somewhat dangerous) to trap and are not numerous. Adult females with cubs-of-the-year may tend to spend more time in the open than other bears, and such family groups can be approximately identified by group size, age of cubs, location, etc. Consequently the only long-term index of abundance for bears in Yellowstone has been an annual "count" of such family groups. The index is quite variable, so it is essential to learn as much about the effect of variability as possible, and to look for ways to improve the index. For further study here, logarithms of the index count are used because a linear relationship would result if the counts are directly proportional to population abundance. A plot of the data (Fig. 3.3) shows the substantial variability.

The jackknife, the bootstrap, and cross-validation can be used to study the index. To use the jackknife approach, one proceeds as in the example shown in Section 3.7. There are 19 annual values of the index, so the original data are copied 19 times, and each of the paired items (year and ln count) is removed in turn and placed at the top of the table. The gaps in the main body of data are then filled by moving the data below up one cell. For each of the paired columns of data, one then estimates a slope (using the SLOPE function in EXCEL) and calculates the intercept from y and x means. This thus gives the basis for a regression line at the bottom of each set of data. This regression line is then used to compute an estimate for the missing point (using the x-value at the top of the table) and that prediction is placed below the value left out located at the top of the table. The resulting 19 data pairs then

provide data for calculation of CVE by eq. (3.6). The first two columns of a calculation appear in a table below.

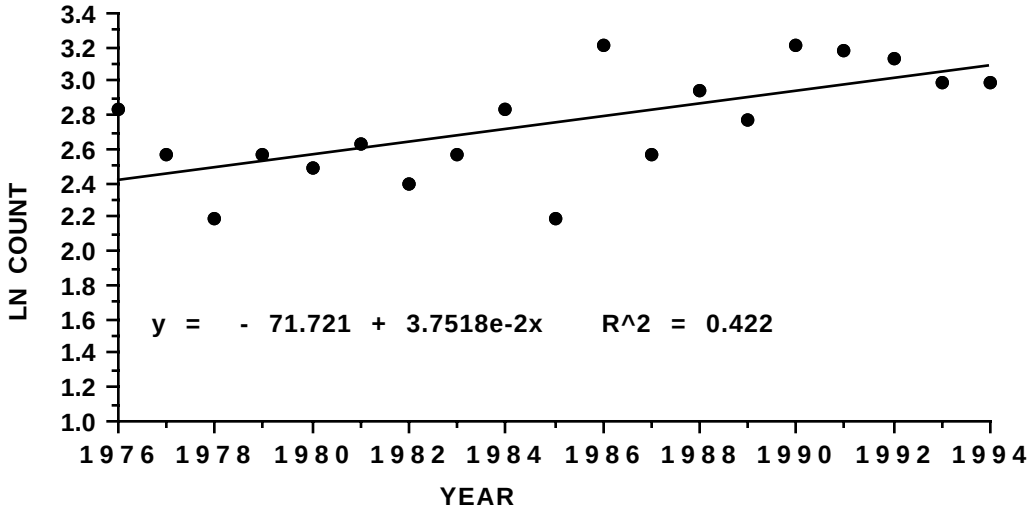

Fig. 3.3 Index of grizzly bear abundance in Yellowstone National Park.

The slope estimates at the bottom of the table are used to jackknife the data for calculating a standard error from eq. (3.11) which turns out to be 0.0104. This is perhaps most useful if divided by the jackknife mean, giving a coefficient of variation of 0.28, indicating the considerable variability in the data. The data can also be used to calculate the jackknife bias estimate of eq.(3.10), which appears to be very small. The operation can be described in steps as follows:

(1) Duplicate the two columns of data (x and y variables) n times, where n is the number of observations available.

(2) Remove each value in turn and put it above the table of values, leaving space for a predicted value.

(3) Move up the data to close the gaps.

(4) Calculate slopes and x and y means for each column.

(5) Use this regression data to calculate a predicted value for the x-value of the item removed from that column and place the predicted value below the removed value. The squared difference is then summed and divided by n to calculate CVE.

(6) Use the calculated slope values to produce jackknife estimates of standard error and bias.

All of the above provides some information on how an index behaves. It is, however, more useful in the situation where we have several possible candidates for an index, as the estimates of bias and CVE (and possibly other statistics calculated for the data) can be used to decide which of the candidates might give the best notion of trends in the bear population, which is of major importance in managing an

important species. An improvement in the index is available by way auxiliary variables that provide a correction for the variation in visibility of bears, which presumably is at least partially responsible for fluctuations in the number seen from year to year. This improved index was described by Eberhardt et al.(1999).

Table of first few columns of data from Jackknifing bear index.

LEDT OUT				2.8332	1	2.5649	2
PREDICT				2.3140		2.4304	
DEV SQ				0.2696		0.0181	
ORIGINAL DATA							
		LN COUNT					
		Y	X	Y	X	Y	X
1976	17	2.8332	1	2.5649	2	2.8332	1
1977	13	2.5649	2	2.1972	3	2.1972	3
1978	9	2.1972	3	2.5649	4	2.5649	4
1979	13	2.5649	4	2.4849	5	2.4849	5
1980	12	2.4849	5	2.6391	6	2.6391	6
1981	14	2.6391	6	2.3979	7	2.3979	7
1982	11	2.3979	7	2.5649	8	2.5649	8
1983	13	2.5649	8	2.8332	9	2.8332	9
1984	17	2.8332	9	2.1972	10	2.1972	10
1985	9	2.1972	10	3.2189	11	3.2189	11
1986	25	3.2189	11	2.5649	12	2.5649	12
1987	13	2.5649	12	2.9444	13	2.9444	13
1988	19	2.9444	13	2.7726	14	2.7726	14
1989	16	2.7726	14	3.2189	15	3.2189	15
1990	25	3.2189	15	3.1781	16	3.1781	16
1991	24	3.1781	16	3.1355	17	3.1355	17
1992	23	3.1355	17	2.9957	18	2.9957	18
1993	20	2.9957	18	2.9957	19	2.9957	19
1994	20	2.9957	19				
MEANS		2.75	10.0	2.75	10.5	2.76	10.4
SLOPES		0.0375		0.0457		0.0394	
INTERCEPTS		2.3776		2.2683		2.3516	
S.S. OF		SLOPES		0.0001		0.0000	

3.8 The Monte Carlo method

In many situations, it is desirable to seek a way to check on the validity of possible estimators. If the stochastic process leading to the data under study can be modelled in a realistic manner, then it is usually possible to test estimation and analysis methods by "Monte Carlo" simulations. Many detailed papers and a sizable number of books deal with such approaches, and all that will be attempted here is to provide a sketch of the method, and a simple example. Exercise 1.16.6 discusses simulation of a continuous frequency distribution, the exponential distribution. The underlying model for survival times is, in fact, the exponential, although survival may also need to be described by more complex models. Given a way to generate a sample from a plausible distribution, one can then use such data to test estimation or analysis schemes.

For a concrete example, we consider the percentile confidence limits discussed in Chapter 2, and demonstrated in Fig. 2.2, and ask whether these

limits are valid. This question is usually discussed in terms of coverage. For convenience, consider 95% confidence limits. These are described as limits that should include the true but unknown mean in 95% of a very large series of repetitions of the same process from which a given observed sample is generated. Note that nothing is said about a particular case -- it is only the long-run average that we can depend on. If confidence limits are properly constructed, then they should "cover" the (unknown) true mean 95% of the time. If we assume that observed survival time data come from an exponential distribution, then we can generate a very large number of samples of n = 20 "observations", calculate bootstrap confidence limits from these samples, and see how well they "cover" the true mean. In this case, we can know the true mean, inasmuch as it can be calculated for the exponential distribution, $E(x) = 1/\beta$. Using $\beta = 0.01$ results in an expected ("true") mean of $1/\beta = 100$.

A BASIC program (Program EXPON SIMUL) was used to study the confidence limits. It turns out that 1,000 simulation runs with 1,000 bootstraps for each sample of n = 20 yields 906 cases where the calculated percentile limits included the true mean of 100, whereas one would expect 950 cases inside the limits for a true 95% level of significance. Note that this result (906 of 1,000) is subject to sampling error; a binomial calculation gives $v(p) = p(1-p)/1000$ where $p = 0.95$, so that two standard errors on p will be about 0.013. Consequently, it would appear that the bootstrap "coverage" is significantly short of the expected 95%. Nonetheless a nominal 91% isn't too bad for confidence limits. Fig. 3.4 provides an example of coverage from this study.

The exponential distribution is sharply skewed and the standard deviation equals the mean so that the survival time thus generated is highly variable. For an alternative, we can run the Monte Carlo study using normally distributed variables with the same mean (100) and a smaller standard deviation (10). This can be done by using the Box-Muller approximation (Bratley et al. 1983) to generate unit normal random variables, replacing the exponential in a BASIC program. This generates two approximately normal random variables with zero mean and variance of unity from two uniform random variables, and these are then transformed to have standard deviation of 10 and mean of 100. Running 1,000 simulations each using 1,000 bootstraps on samples of n = 20 from the normal distribution gives coverage of 931, appreciably closer to the expected 95%.

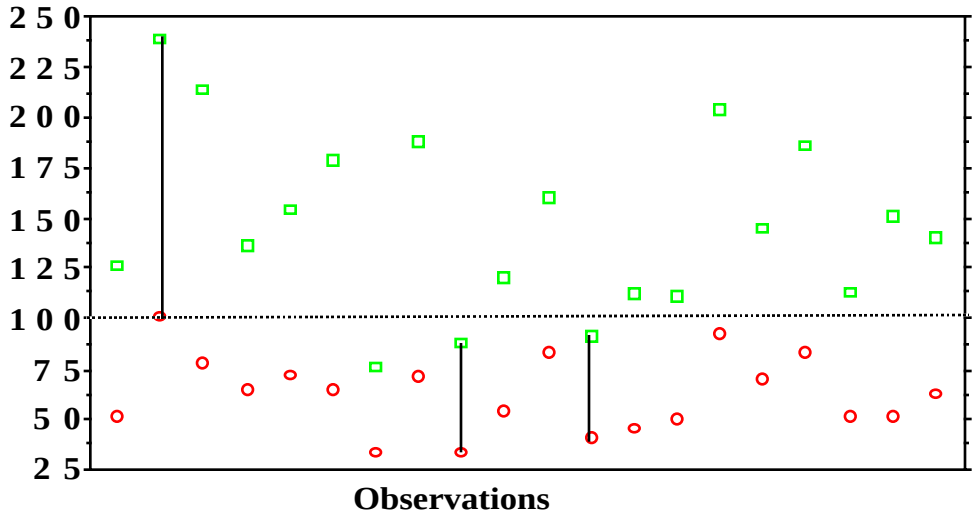

Fig. 3.4. An example of "coverage" for the simulation to test bootstrap percentile limits on simulated data from an exponential distribution. The figure shows confidence limits for a sample of 20 observations out of the 1,000 simulations used to test confidence limits calculations (with 1,000 bootstraps per sample of 20 observations). The dotted line shows the position of the true mean of 100, while squares represent upper 95% confidence limits and circles the lower limits. Lines connect the limits for the three cases where the limits did not include the true mean. In this sample, coverage was 17/20 = 0.85.

Inasmuch as we are dealing with means, the usual approach to confidence limits would be to calculate a variance from the original data, and obtain confidence limits with a multiplier from the t-distribution. Such results can easily be simulated, using the same methods for generating exponential and normal random variables. Program T SIMUL NORM was used to simulate samples and confidence limits from the normal distribution, and the same program was used with the exponential generator. Results using sample sizes of n = 10, 20, and 30 appear in the following table. These results suggest that constructing confidence limits in the usual manner from exponential data does a little better job than bootstrapping, and for data from a normal distribution the limits are within sampling error of the expected 95%, while bootstrapping falls a little short. One would not, of course, use bootstrapping to obtain confidence limits on means. It is best reserved for situations where there is no convenient theoretical approach.

Sample	Exponential simulation		Normal simulation	
size	Bootstrapping	Usual limits	Bootstrapping	Usual limits
10	837	900	904	944
20	906	923	931	941
30	927	921	933	953

3.9 The delta method

The delta method is a useful adjunct to bootstrapping. It has been used for many years to approximate the variance of complex functions of random variables. It is obtained from a Taylor expansion of the function in which the

second degree terms are retained, and rearranged to represent variances of the random variables. The expression is as follows:

$$V[g(x)] = \sum_{i=1}^{n} V(x_i)\left(\frac{\partial g}{\partial x_i}\right)^2 + 2\sum_{i<j} \mathrm{cov}(x_i x_j)\left(\frac{\partial g}{\partial x_i}\right)\left(\frac{\partial g}{\partial x_j}\right) \qquad (3.12)$$

where V[g(x)] represents the variance of some function, g(**x**), where **x** is a vector of random variables, $x_1, x_2, \ldots, x_n$. $V(x_i)$ denotes the variance of the variable x_i, which is multiplied by the square of the partial derivative of $g(x_i)$. Covariance terms are calculated for those cases where i<j. In many cases, it may be that the random variables are independent, so that the covariance terms can be assumed to be zero, and the right-hand portion of eq.(3.12) can then be dropped.

Bootstrapping can be used to calculate a variance for g(**x**) without any need to calculate variances and covariance of the individual random variables or to obtain partial derivatives. The delta method becomes a valuable adjunct, however, when it is possible to design the study in order to minimize V[g(**x**)]. In practice, V[g(**x**)] may be appreciably larger than is desirable, and we may wish to design a new study with larger samples (or to supplement the existing samples). In this case, it is essential to be able to determine the effect of increasing the sample sizes for individual random variables. The delta method provides a way to calculate the effects of changing sample sizes on the overall variance.

Example 3.5 Application of bootstrapping to a complex function.
Obtaining a variance for the Lotka-Leslie model provides a good example of the utility of bootstrapping. The underlying equation for this model is:

$$1 = \sum \lambda^{-x}\, l_x\, m_x \qquad (3.13)$$

Here, λ represents the rate of change of an age-structured population having age-specific survivorship rates l_x and age-specific reproductive rates m_x. The general model for the Lotka-Leslie function does not have a "closed-form" solution. That is, there is no way to write eq.(3.13) in a linear form, that is to provide an expression stated as $\lambda = g(x)$. It is thus necessary to solve eq.(3.13) for λ by an iteritive procedure, i.e., by varying values of λ until one satisfies the equation. Because there is no linear expression for a solution for λ, developing an expression for the variance becomes very difficult. Bootstrapping then provides a convenient approach. One only needs to set up the data on l_x and m_x in tables, sample these tables of data with replacement, and calculate values of λ from the samples. The percentile method then provides convenient confidence limits.

In many instances, the samples available for calculations are too small to make calculations from eq.(3.13) feasible. An alternative may then be needed. A useful approximation (Eberhardt 1985) is:

$$\lambda^a - s\lambda^{a-1} - l_a m \left[1 - \left(\frac{s}{\lambda}\right)^{w-a-1}\right] = 0 \qquad (3.14)$$

Here, a is the age at which full reproductive rate is achieved, l_a is survival from birth to age a, s is survival beyond that age, and w is an age at which calculations are truncated in order to compensate for the effects of senility. This equation again must be solved by iteration, and can readily be bootstrapped. The delta method can be used to study the components of variance and thus to determine the effect of increasing sample sizes for the several components on the variance estimate for λ. In several examples, the delta method gives very much the same variance estimate as bootstrapping. Another benefit of the delta method calculations is that the partial derivatives serve to indicate the relative importance of the several components, indicating, for example, that small changes in adult survival have the maximum effect on λ. Because there is no linear solution for λ, the delta method has to be applied by using implicit differentiation. Solutions appropriate for eq.(3.14) appear in the following references, which also give details and result of the application of bootstrapping to this complex function. Calculations for grizzly bears appear in Eberhardt et al. (1994), for sea otters in Eberhardt (1995), for monk seals in Gilmartin and Eberhardt (1995), and for manatees in Eberhardt and O'Shea (1995). Selected examples appear in Chapter 11.

3.10 Exercises

3.10.1 Inasmuch as bootstrapping is a sampling procedure, additional runs of B bootstraps will give slightly different results, even if B is large. Conduct a bootstrapping check on the data of Example 3.1 to see how your bias adjustment compares with the results given there. Use B=2000. Do 10 trials and record results on a summary sheet (don't forget to use PASTE SPECIAL and VALUES or you may get a statement like "Circular References" or "Link to another spreadsheet"). This should show that the bias is consistent, and that the corrected value is a much better estimate of the true value. However, when there is an unbiased estimate based on theory (as in this case), one obviously should use that value. The bias estimate is important only when you don't have an estimate that is known to be unbiased (which is often the case with ecological data, even though it might not be a widely recognized fact). Also number the observations serially (1-10) and calculate the correlation coefficients.

3.10.2 Bias corrections.
Use the data of Example 2.3 (calculations in Exercise 2.10.4) to further explore bias corrections. In that example we used regressions of the natural logarithm of number of survivors (Fig. 2.5) on year to estimate a survival rate (slope of the regression line) and then transformed it back to an annual rate by calculating y=exp(b). Bootstrap confidence limits were obtained and also transformed back to annual rates. Use eq.(3.2) to examine the bias in transforming back. When there is an evident bias, one should examine the confidence interval on the estimate to see if the bias is large relative to the confidence interval.

$$\widehat{\text{bias}}_B = \hat{\theta}^*(\bullet) - t(\hat{F}) \qquad (3.2)$$

3.10.3 Make a frequency distribution of z (eq.(2.9)) using the correlation coefficients computed in exercise 3.10.1. Does this look like a normal distribution as assumed in calculating confidence limits under the usual theory? Compare your results with Exercise 2.10.5.

3.10.4 The regression bootstrap of Example 2.2 used parametric bootstrapping in which deviations from a model fitted to the original data are bootstrapped. In exercise 2.10.7 we tried bootstrapping the x,y pairs directly ("nonparametric" bootstrapping) and got some strange-looking results. However, larger samples (more x and y values) appear to give results comparable to parametric bootstrapping. Efron and Tibishirani warn that the parametric approach is "model-dependent", i.e., if the model is wrong, the results may be doubtfully useful. Hence, its worthwhile to repeat the exercise using the data of Example 3.3. Doing this directly is cumbersome, so it is best to use the program furnished in the Appendix. Compare your results with the slope and confidence intervals given by the regression program in EXCEL. This exercise is worthwhile in that ecologists use regressions with smallish samples and the independent variables are not always known with certainty. There don't seem to be any guidelines as to sample sizes in such cases, so its wise to use both parametric and nonparametric approaches and to check for bias (Eq. (3.2)) if you want to be comfortable with your results. The frequency diagrams of Exercise 2.10.7 were distinctly bimodal, making it clear that the nonparametric approach is not advisable with only 10 pairs of observations.

3.10. 5 Example 3.4 gives the approach to jackknifing a regression line in which logarithms of data on an index of bear abundance are fitted by linear regression (Fig. 3.3) and the fit examined by cross-validation, with a check on bias from eq.3.10. Complete the analysis just as in Example 3.4. Compute the cross-validation error (CVE), jackknife standard error of the slopes, and Biasjack of the slopes. Compare the jackknife standard error with that of the slope computed with the usual regression analysis (given in EXCEL tools menu). Also compare CVE with the residual mean square of the regression analysis.

3.10. 6 Jackknifing was used in Example 3.10.5 because it is fairly easy to apply and we could compute CVE of eq. (3.6) in the same operation. However, bootstrapping has some advantages, and likely should be used to estimate bias and confidence limits whenever it is feasible. Use the data of Example 3.10.5 to conduct parametric bootstrapping to compute the bootstrap bias estimate of Eq. (3.2) and 95% confidence limits on the slope. Use 1,000 bootstraps and parametric regression bootstrapping (for convenience in calculations using EXCEL – with 19 observations, I would be inclined to try both parametric and non-parametric approaches). The nonparametric regression bootstrap can readily and quickly be computed from a program in the Appendix.

3.10.7 The approach of Section 3.5 is most readily calculated by using a programming language. However, it is feasible to do the calculations in EXCEL if one is willing to devote several hours to the job. A program in the Appendix will do the job in short order and should be used to repeat the results of Section 3.5.